DOCUMENT RESUME

ED 167 598 TM 008 371

AUTHOR	Goldberg, Zandra S.; Spartz, James L.
TITLE	The Use of Longitudinal Analysis to Identify More and Less Effective Schools.
PUB DATE	Mar 78
NOTE	9p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Resources; Elementary Education; Expenditures; Family Income; Family Influence; Multiple Regression Analysis; *Predictor Variables; *Program Effectiveness; School Policy; *School Support; Socioeconomic Background; Success Factors
IDENTIFIERS	Delaware; *School Evaluation

ABSTRACT

Three longitudinal studies of student achievement test data were conducted to examine the extent to which the variation in student achievement can be explained by differences in student background (including socioeconomic and prior achievement data) and school resources. This model assumes that achievement test performance may be predicted by these factors plus the effectiveness of the school program. Data were analyzed by school, and included means of aptitude and achievement test scores, student background data, and school resource information. Students were tested in first, fourth, and eighth grades. One study compared student achievement from fourth to eighth grade, while two studies compared achievement from first to fourth grade. It was concluded that longitudinal analysis of these data can identify schools whose achievement appears to be above or below expectancy, but the results of one study are not reliable so it is essential that a replication study be conducted. However, where the data for the study included data from both elementary schools and middle schools, this method was not successful in identifying effective schools. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167598

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE USE OF LONGITUDINAL ANALYSIS TO IDENTIFY MORE AND LESS EFFECTIVE SCHOOLS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Zandra S. Goldberg
Del. Dept. Publ. Inst.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Zandra S. Goldberg

James L. Spartz

Delaware Department of Public Instruction

Paper presented at the 1978
Convention of the
American Educational Research Association
Toronto, Canada

TM008 371

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the educators and students in Delaware who participated in the Delaware Educational Assessment Program (DEAP). In addition, we would like to acknowledge the DEAP staff members which presently include Chester W. Freed, Alice L. Valdes, Janet Wall, and Wilmer E. Wise, Director. Special thanks are also extended to Diana Robertson for her assistance in the tabulation and analysis of the data and Julia Burris for typing this manuscript.

INTRODUCTION

Using data collected as part of the Delaware Educational Assessment Program (DEAP), three longitudinal studies of student achievement test data were conducted. These studies have in common the following focus:

> The examination of the extent to which the variation in student achievement can be explained by differences in student background (including socioeconomic and prior achievement data) and school resources.

Specifically, it was our desire to identify schools whose mean achievement was atypically high or low after considering student background and school resources. The further study of these schools may then lead to the determination of in-school characteristics which differentiate between more and less effective schools.

The model employed in these studies assumes that achievement test performance of a group of students in a school at some point in time is a function of the following four factors:

(1) their achievement at some previous point in time;

(2) their family background;

(3) the resources expended by the school; and

(4) the effectiveness of the school program.

Students' achievement test performance, aggregated at the school level, may be adjusted for previous achievement test performance, family background, and school resources. If differences between schools' adjusted achievement test means are observed, these differences, to some extent, may be attributed to differences in the schools' educational program. In this manner, the relative effectiveness of school programs within the state may be evaluated.

METHOD

Three longitudinal studies were conducted. Each study used prior mean achievement, student background, and school resource information to predict school mean achievement.

Subjects

From 1972 to 1976, all Delaware first-, fourth-, and eighth-grade students participated in DEAP. Subjects for each study were those students who remained in the same school district across the time span of each study. Table I presents data which describe the sample for each study.

TABLE 1
DESCRIPTION OF THE SAMPLE FOR THE THREE LONGITUDINAL STUDIES

Study Number	Study Dates	Grades Involved	Number of Schools	Number of Students Matched	Percent of Total Tested Matched
1	1972 - 1975	one to four	103	5,667	61.3
2	1972 - 1976	four to eight	41	9,863	66.3
3	1973 - 1976	one to four	99	5,370	65.2

Materials

The DEAP test batteries include reading, listening, and mathematics tests for grade one. At grades four and eight, the DEAP test batteries include verbal and quantitative aptitude tests, reading, English, mathematics, science, and social studies achievement tests. These tests range from 30 to 100 items each. Several tests are exact reprints of standardized tests while other tests were developed specifically for Delaware. Most tests contain a mixture of standardized items and items developed specifically for Delaware. For each test, students' scores were transformed into T-scores.

Student background data were obtained from three sources. Each year school principals provided estimates of such variables as the average education level of parents. State level records were accessed yearly to obtain values on variables such as student dropout rate and the full value of real estate per pupil. Additional student background data were obtained from U. S. Census results (e.g. median monthly rent). Unlike student background information, school resource information was obtained from one source. Information on such variables as the Number of Teachers per 1,000 Pupils and the Percent Master's Degree Teachers were obtained solely from state level files. Each school's value for each student background and school resource variables was the mean of the values assigned that school each year of the study.

Procedure

In each study, the data were aggregated at the school level; that is, the school was the unit of analysis. The predictor variables, consisting of indicators of initial school mean achievement, student background, and school resources, were factor analyzed (varimax rotation) and reduced

into nine factors. Estimated factor scores for each school were used to predict (via stepwise regression analysis) school mean achievement test performance in the latter year of the study. The regression analyses were stopped prior to the step in which the addition of a new variable failed to increase the amount of variance accounted for by ten percent. In each study, the regression equation which best predicted the criterion was employed. Large residuals between predicted and actual mean achievement indicate atypically high (positive residual) and low (negative residual) school mean achievement.

RESULTS

In each study regression analysis was employed to predict mean achievement test performance.

Comparison of Study 1 and 3

Studies 1 and 3 were similar in that each predicted grade four achievement using grade one achievement, student background, and school resource information. The difference between the two studies was the years of data collected (Study 1 - 1972 through 1975, and Study 3 - 1973 through 1976). The correlation coefficients between the criterion tests and the predictor variables observed in Study 1 replicated in Study 3. Although not identical, similar factor structures were evident in Studies 1 and 3. The regression equation employed in Study 1 was similar to that derived in Study 3; both regression equations were primarily composed of student prior achievement and socioeconomic status information. However, a substantial number of the significant test residuals observed in Study 1 did not replicate in Study 3. This lack of replicability was observed regardless of the significance level employed. That is, using one standard error of estimate (S.E.), 60 percent did not replicate; at 1.4

S.E., 59 percent did not replicate, and at 1.8 S.E., 69 percent did not replicate. Thus, it was concluded that a conservative significance level is not adequate protection from incorrectly isolating extreme cases.

Comparison of Studies 1 and 3 with 2

Although all three studies were similar in design, Studies 1 and 3 predicted grade four achievement while Study 2 predicted grade eight achievement. In the first- to fourth-grade studies, a substantial portion of the differences between schools on the fourth-grade tests remained unexplained after considering first-grade achievement, student background, and school resources. Regression analyses were not able to explain from 23 to 48 percent of the variance (R^2 ranged from 52 to 77). This unexplained variance, at least in part, may be attributed to in-school characteristics. In the fourth- to eighth-grade study, most of the variation between schools on the eighth-grade achievement tests can be accounted for by the input variables (see Table 2). Thus, there do not appear to be eighth-grade schools achieving substantially above or below expectancy. Study 3's results do not facilitate identification of more or less effective school programs within the state.

TABLE 2

PERCENT OF VARIANCE ACCOUNTED FOR IN EACH LONGITUDINAL STUDY (R^2)

Criterion Test	One to Four 1972 - 1975	Four to Eight 1972 - 1976	One to Four 1973 - 1976
Reading	.77	.92	.72
English	.73	.92	.52
Mathematics	.72	.90	.70
Science	.65	.84	.64
Social Studies	.76	.91	.61

CONCLUSION

The integration of the results of the three longitudinal studies provides the following two conclusions regarding school program evaluation. The first conclusion is based on the replicability of the elementary school study results. Longitudinal analysis of elementary school data can identify schools whose achievement appears to be above or below expectancy. However, the results of one study are not reliable indicators of school effectiveness. It is, therefore, essential that a replication study be conducted before final determination of schools achieving above or below expectancy. Those schools in which significant test residuals replicated may now be the focus of further study.

Further research into the causes of the lack of replicability between Studies 1 and 3 is needed. That is, the inconsistent results may be attributed to school level achievement and/or socioeconomic levels changing differently during the time span of Study 1 as compared to Study 3. The lack of replicability may also be the result of the statistical procedures employed. Specifically, the differences between Study 1 and 3's residuals may be an artifact of the specific regression equations used.

The second conclusion is based on the comparison of the elementary school studies with the middle school study. Unlike the elementary school data, longitudinal analysis of middle school data was not successful in identifying exceptional schools. Knowledge of student prior achievement, student background, and school resources yielded very accurate predictions. It is likely that the school programs studies were equally effective relative to the assessment instruments used.

ZSC:jb

3/3/78